Digital Curation as a Key Component in Research Infrastructures: From Data Preservation to Processes Preservation and Verification

© Andreas Rauber

Institute of Software Technology and Interactive Systems
Vienna, Austria
rauber@ifs.tuwien.ac.at

With the advent of data-driven science, also referred to as, for example, the Fourth Paradigm, Big Data, and other similar concepts, the need to safeguard the investments made into collecting and preparing massive amounts of data (some of which is unrecoverable) has drastically gained importance. Providing digital preservation of research data is thus emerging as a service that has to be provided by sophisticated research infrastructure frameworks. Yet, with the complexity of research processes increasing, the needs for preservation stretch beyond merely maintaining data accessible. Capturing and documenting the context of its creation and use is an enormous task, requiring sophisticated representation information networks. Even more challenging, complex processes are an integral part of data provenance. We thus also need to capture, preserve, and maintain usable a series of data processing routines and modules in order to be able to establish the validity of scientific analysis, to repeat earlier computations on new data, in short to make full use of the opportunities offered by data-intensive science.

This tutorial will start with a brief review of the classical challenges in digital preservation. It will then move on to motivate the need for process preservation as part of data curation. This will be followed by a presentation of approaches to facilitate process preservation, most notably process context capture as well as recommendations on how to ease process preservation by proper design.

We will also address legal arrangements to counter loss of proprietary information required for maintaining processes executable. Last, but not least we will discuss a framework for evaluating processes to verify authentic behavior upon re-execution, identifying information to be captured and processing steps to be performed upon process design and preparation for preservation.

Proceedings of the 14th All-Russian Conference "Digital Libraries: Advanced Methods and Technologies, Digital Collections" — RCDL-2012, Pereslavl-Zalesskii, Russia, October 15-18 2012.